Nov 1979
in
Interlinking
of Computer Networks
K. G. Beauchamp (ed)
p 287-305
© 1979
D. Reidel Publishing
Dordrecht, Holland.

DARPA ACTIVITIES IN PACKET NETWORK INTERCONNECTION

Vinton G. Cerf

U.S. Department of Defense, Advanced Research Projects Agency

DARPA Activities in Packet·Network Interconnection

1. The military requirement for Network Interconnection Technology

A fundamental premise of all current Command, Control and Communications (C3) research is that digital technology and computing systems will play a central role in the future. It is already apparent that computers are being employed in tactical as well as strategic military equipment both to replace older technology and to significantly improve our ability to control weapons systems. Automatic fire control, computer-aided information collection and correlation, computer-controlled array sensors and computer-controlled communication systems are all symptomatic of an increasingly automated C3 environment.

To make this collection of computers, sensors, and databases useful, it is crucial that the components be able to intercommunicate. In the late 1960's, DARPA began a systematic research effort into technologies which would support computer communication in both a strategic and a tactical environment. The concept of "packet switching," in which the unit of switching was a "packet" containing a few tens or hundreds of characters of information, was developed. This idea was in direct contrast to the notion of setting up and clearing down circuits from point-to-point. The packet switching idea paved the way for highly efficient sharing of broadband transmission resources among physically dispersed and very "bursty" traffic sources. The new packet switching concept was perfectly suited to the high peak-to-average rate communication characteristics of computer systems.

Out of this research, the packet switched ARPANET (1) was developed. Based on store-and-forward concepts but using high speed telephone circuits, short packets, and fast minicomputers, the ARPANET demonstrated that short response times on the order of 200 milliseconds could be achieved while also supporting high bandwidth transmission when needed.

For all its utility, however, the ARPANET did not serve as a good model for the support of tactical mobile or ocean-going computer communication. DARPA developed two other concepts, packet radio (2,3) and packet satellite (4), to deal with these unique requirements. Both systems use wideband shared radio channels. The packet radio system operates in a 20 MHz spread spectrum mode, transmitting data at either 100 kb/s or 400 kb/s. Packet radios have built-in microprocessors which act as store-and-forward devices, and also effect control over access to the shared radio channel. The system is fully mobile and self-organizing.

The packet satellite system supports four satellite ground stations each of which is equipped with a minicomputer which controls access to the shared satellite channel. Identical software in each ground station permits fully distributed control of the satellite network and makes it possible to capitalize on the broadcast nature of the single, shared satellite transponder channel. Shared channel bandwidths of 64 Kb/s have been used and will be increased to 1.5 Mb/s in the future. The system is designed to handle tens to hundreds of ground stations.

In a related development called "Ethernet," Xerox Palo Alto Research Center introduced the concept of a share co-axial cable (5). Hundreds of computers in a single building complex connected to a common co-axial cable. Each computer transmission is heard by all others attached to the cable, as in the satellite case. Computers transmit randomly, at will, but quickly detect if they have "collided" with another transmitter. When such a collision is detected, those transmitters in collision stop and each delay for a random interval before retrying. The system runs at about 3Mb/s and supports and effective utilization of about 30%. Since bandwidth is very easy to obtain in this system, low utilization is perfectly acceptable.

These four examples, ARPANET, Packet Radio, Packet Satellite, and Ethernet illustrate the notion that different communication media are needed for different applications. Ethernet ideas might serve well in garrison or aboard a ship. Packet radio concepts are crucial for local area mobile communication (e.g., land mobile, ground-air, ship-ship). ARPANET technology is appropriate for fixed installations such as in CONUS or Europe. Finally, packet satellite supports wide geographic coverage while permitting efficient and dynamic allocation of transmission capacity as needed. The conclusion is that many different transmission technologies are needed for military operations and therefore, a sensible C3 system must incorporate a strategy for the interoperation of dissimilar computer communication networks.

This requirement for interoperability leads naturally to the concept of packet network interconnection. Techniques and options for achieving network interconnection have been explored by a number of researchers over the last few years (6-13). The basic concept is to connect different networks by means of gateway computers. Traffic from a computer on one net to a computer on another flows through intermediate networks and gateways to the destination. The mechanisms for moving the packets among the nets allow for alternate routing through redundant gateways in the case of failure. In addition to this, the communication protocols (formats and conventions) used by the sources and sinks of traffic are highly robust and can deal with loss, duplication and disordering of packets as they flow from one net to another.

The general picture is a richly interconnected collection of networks of differing internal operation with common higher level protocols and internetwork gateways to bind the system together in a robust and survivable fashion.

This interconnection concept has one other crucial advantage for military operation. Because it is designed to support effective communicaton between computers on dissimilar networks, it can also serve to support evolutionary transitions from one networking technology to another. As new networks are formed, these can grow while the old ones shrink. Interoperability can be maintained through gateways and common protocols until the old networks are phased out.

Thus, the problems of dealing with dissimilar tactical and strategic nets and with evolving computer communication network technology can be solved in a single stroke.

DARPA is currently engaged in an ambitious program which is binding together its internetting technology, integrated data, voice, graphics capabilities, secure networking and opeating system technology, and distributed database technology into a coherent framewrok upon which a highly flexible and long-lived C3 system could be constructed.

2. The Catenet Model for Internetworking

The term "catenet" was introduced by L. Pouzin in 1974 in his early paper on packet network interconnection (6). The U.S. DARPA research project on this subject has adopted the term to mean roughly "the collection of packet networks which are connected together." This is, however, not a sufficiently explicit definition to determine, for instance, whether a new network is in conformance with the rules for network interconnection which make the catenet function as confederation of co-operating networks. This section attempts to define the objectives and limitations of the ARPA-internetworking project and to make explicit the catenet model on which the internetworking strategy is based.

Objectives

The basic objective of this project is to establish a model and a set of rules which will allow data networks of widely varying internal operation to be interconnected, permitting users to access remote resources and to permit intercomputer communication across the connected networks.

One motivation for this objective is to permit the internal technology of a data network to be optimized for local operation but also permit these locally optimized nets to be readily interconnected into an organized catenet. The term "local" is used in a loose sense, here, since it means "peculiar to the particular network" rather than "a network of limited geographic extent." A satellite-based network such as the ARPA packet satellite network therefore has "local" characteristics (e.g., broadcast operation) even though it spans many thousands of square miles geograpically speaking.

A second motivation is to allow new networking technology to be introduced into the existing catenet while remaining functionally compatible with existing systems. This allows for the phased introduction of new and obsolescence of old networks without requiring a global simultaneous change.

Assumptions

One of the first questions which must be settled in a project of this sort is "what types of data networks should be included in the catenet model?" The answer to this question is rooted in the basic functionality of each candidate network. Each network is assumed to support the attachment of a collection of programmable computers. Our essential assumption is that any participating data network can carry a datagram (14,15) containing no less than

1000 bits of data not including a local network header containing local control information. Furthermore, it is assumed that the participating network allows switched access so that any source computer can quickly enter datagrams for successive and different destination computers with little or no delay (i.e., on the order of tens of milliseconds or less switching time).

Under these assumptions, we can readily include networks which offer "datagram" interfaces to subscribing host computers. That is, the switching is done by the network based on a destination address contained in each datagram passing across the host to network interface.

The assumptions do not rule out virtual circuit interface networks, nor do they rule out very fast digital circuit switching networks. In these cases, the important functionality is still that a datagram can be carried over a real or virtual circuit from source to destination computer, and that the switching delay is below a few tens of milliseconds.

An important administrative assumption is that the format of an internet datagram can be commonly agreed, along with a common internet addressing plan. The basic assumption regarding datagram transport within any particular network is that the datagram will be carried, embedded in one or more packets, or frames, across the network. If fragmentation and reassembly of datagrams occurs within a network it is invisible for purposes of the catenet model. Provision is also made in the datagram format for the fragmentation of datagrams into smaller, but identically structured datagrams which can be carried independently across any particular network. No a priori position is taken regarding the choice between internal (invisible) fragmentation and reassembly or external (visible) fragmentation. This is left to each network to decide. We will return to the topic of datagram format and addressing later.

It is very important to note that it is explicitly assumed that datagrams are not necessarily kept in the same sequence on exiting a network as when they entered. Furthermore, it is assumed that datagrams may be lost or even duplicated within the network. It is left up to higher level protocols in the catenet model to recover from any problems these assumptions may introduce. These assumptions do not rule out data networks which happen to keep datagrams in sequence.

It is also assumed that networks are interconnected to each other by means of a logical "gateway" (16). As the definition of the gateway concept unfolds, we will see that certain types of network interconnections are "invisible" with respect to the catenet model. All gateways which are visible to the catenet model have the characteristic that they can interpret the address fields of internet datagrams so as to route them to other gateways or to destinations within the networks directly attached to (or associated with) the gateway. To send a datagram to a destination, a gateway may have to map an internet address into a local network address and embed the datagram in one or more local network packets before injecting it into the local network for transport.

The set of catenet gateways are assumed to exchange with each other at least a certain minimum amount of information to enable routing decisions to be made, to isolate failures and identify errors, and to exercise internet flow and congestion control. Furthermore, it is assumed that each catenet gateway can report a certain minimum amount of status information to an internetwork monitoring center for the purpose of identifying and isolating catenet failures, collecting minimal performance statistics and so on.

A subset of catenet gateways may provide access control enforcement services. It is assumed that a common access control enforcement mechanism is present in any catenet gateway which provides this service. This does not rule out local access control imposed by a particular network. But to provide globally consistent access control, commonality of mechanism is essential.

Access control is defined, at the catenet gateway, to mean "permitting traffic to enter or leave a particular network." The criteria by which entrance and exit permission are decided are the responsibility of network "access controllers" which establish access control policy. It is assumed that catenet gateways simply enforce the policy of the access controllers.

The Catenet Model

It is now possible to offer a basic catenet model of operation. Figure 1 illustrates the main components of the model. Hosts are computers which are attached to data networks. The host/network interfaces are assumed to be unique to each network. Thus, no assumptions about common network interfaces are made. A host may be connected to more than one network and it may have more than one connection to the same network, for reliability.

Gateways are shown as if they were composed of two or more "halves." Each half-gateway has two interfaces:

1. An interface to a local network.

2. An interface to another gateway-half.

One example is given of a gateway with three "halves" connecting networks A, B, and C. For modelling purposes, it is appropriate to treat this case as three pairs of gateway halves, each pair bilaterally joining a pair of networks.

The model does not rule out the implementation of monolithic gateways joining two or more nets, but all gateway functions and interactions are defined as if the gateways consisted of halves, each of which is associated with a specific network.

A very important aspect of this model is that no a priori distinction is made between a host/network interface and a gateway/network interface. Such distinctions are not ruled out, but they are not relevant to the basic catenet model.

As a consequence, the difference between a host which is connected to two networks and a monolithic gateway between networks is entirely a matter of whether table entries in other gateways identify the host as a gateway, and whether the standard gateway functionality exists in the host. If no other gateway or host recognizes the dual net host as a gateway or if the host cannot pass datagrams transparently from one net to the next, then it is not considered a catenet gateway.

The model does not rule out the possibility of implementing a gateway-half entirely as part of a network switching node (e.g., as software in an ARPANET IMP). The important aspect of gateway-halves is the procedure and protocol by which the half-gateways exchange datagrams and control information.

The physical interface between directly connected gateway halves is of no special importance. For monolithic gateways, it is typically shared memory or an interprocess communication mechanism of some kind; for distinct gateway halves, it might be HDLC, VDH, any other line control procedure, or inter-computer buss mechanism.

Hidden Gateways

No explicit network hierarchy is assumed in this model. Every network is known to all catenet gateways and each catenet gateway knows how to route internet datagrams so they will eventually reach a gateway connected to the destination network.

The absence of an explicit hierarchical structure means that some network substructures may be hidden from the view of the catenet gateways. If a network is composed of a hierarchy of internal networks connected together with gateways, these "hidden gateways" will not be visible to the catenet gateways unless the internal networks are assigned global network addresses and their interconnecting gateways co-operate in the global routing and network flow control procedures.

Figure 2 illustrates a simple network hierarchy. For purposes of, identification, the three catenet gateways have been labelled G(AX), G(BX) and G(CX) to indicate that these gateways join networks A and X, B and X and C and X, respectively. Only G(AX), G(BX), and G(CX) are considered catenet gateways. Thus they each are aware of networks A, B, C and X and they each exchange routing and flow-control information in a uniform way between directly connected halves.

Network X is composed of three internal networks labelled u, v and w. To distinguish them from the catenet gateways, the "hidden gateways" of net X are labelled HG(nm) where "nm" indicate which nets the hidden gateways join. For example, HG(vw) joins nets v and w. The notation for HG is symmetric, ie., HG(vw)=HG(wv).

Gateways G(AX), G(BX), G(CX) exchange connectivity and other flow control information among themselves, via network X. To do this, each gateway half must know an address, local to network X, which will allow network X to route datagrams from G(AX) to G(BX), for example.

From the figure, it is plain that G(BX) is really a host on network B and network v. But network v is not one of the globally recognized networks. Furthermore, traffic from G(AX) to G(BX) may travel from net u to net v or via nets u and w to net v. To maintain the fiction of a uniform network X, the gateway halves of G(AX), G(BX) and G(CX) attached to net X must be aware of the appropriate address strings to use to cause traffic to be routed to each catenet gateway on net X. In the next section, we outline a basic internet addressing philosophy which permits such configurations to work.

Local Gateways

Another element of the catenet model is a "local gateway" associated with each host. The local gateway is capable of reassembling fragmented internet datagrams, if necessary, and is responsible for encapsulation of internet datagrams in local network packets. The local gateway also selects internet gateways through which to route internet traffic, and responds to routing and flow control advice from the local network and attached catenet gateways.

For example, a local gateway might encapsulate and send an internet datagram to a particular gateway on its way to a distant network. The catenet gateway might forward the packet to another gateway and send an advisory message to the local gateway recommending a change in its catenet gateway routing table. Local gateways do not participate in the general routing algorithm executed among the catenet gateways.

Internet Addressing

The basic internet datagram format (17) is shown in Figure 3. By assumption, every network in the catenet which is recognized by the catenet gateways has a unique network number. Every host in each network is identified by a 24 bit address which is prefixed by the network number. The same host may have several addresses depending on how many nets it is connected to or how many network access lines connect it to a particular network.

For the present, it is assumed that internet addresses have the form: Net.Host. "Net" is an 8 bit network number. "Host" is a 24 bit string identifying a host on the "Net," which can be understood by catenet and possibly hidden gateways.

The catenet gateways maintain tables which allow internet addresses to be mapped into local net addresses. Local gateways do likewise, at least to the extent of mapping an "out-of-network" address into the local net address of a catenet gateway.

In general, catenet gateways maintain a table entry for each "Net" which indicates to which gateway(s) datagrams destined for that net should be sent. For each "Net" to which the gateway is attached, the gateway maintains tables, if necessary, to permit mapping from internet host addresses to local net host addresses. The typical case is that a gateway half is connected to only one network and therefore only needs to maintain local address information for a single network.

It is assumed that each network has its own locally specific addressing conventions. To simplify the translation from internet address to local address, it is advantageous, if possible, to simply concatenate a network identifier with the local "host" addresses to create an internet address. This strategy makes it potentially trivial to translate from internet to local net addresses.

More elaborate translations are possible. For example, in the case of a network with a "hidden" infrastructure, the "host" portion of the internet address could include additional structure which is understood only by catenet or hidden gateways attached to that net.

In order to limit the overhead of address fields in the header, it was decided to restrict the maximum length of the host portion of the internet address to 24 bits. The possibility of true, variable-length addressing was seriously considered. At one point, it appeared that addresses might be as long as 120 bits each for source and destination. The overhead in the higher level protocols for maintaining tables capable of dealing with the maximum possible address sizes was considered excessive.

For all the networks presently expected to be a part of the experiment, 24 bit host addresses are sufficient, even in cases where a transformation other than the trivial concatenation of local host address with network address is needed to form the 32 bit internet host address.

One of the major arguments in favor of variable length "addressing" is to support what is called "source-routing." The structure of the information in the "address" really identifies a route (e.g., through a particular sequence of networks and gateways). Such a capability could support ad hoc network interconnections in which a host on two nets could serve as a private gateway. Though it would not participate in catenet routing or flow control procedures, any host which knows of this private gateway could send "source-routed" internet datagrams to that host.

To support experiments with source routing, the internet datagram includes a special option which allows a source to specify a route. The option format is illustrated in Figure 4. The option code identifies the option and the length determines its extent. The pointer field indicates which intermediate destination address should be reached next in the source-selected route.

Source routing can be used to allow ad hoc network interconnections to occur before a new net has been assigned a global network identifier.

In general, catenet gateways can only interpret internet addresses of the form Net.Host. Private gateways could interpret other, local addresses for desired destinations. If a source knew the local addresses of each intermediate private gateway, it could construct a source-route which is the concatenation of the local addresses of each intermediate host.

Local and internet addresses could be inter-mixed in a single source route as long as catenet gateways only had to interpret full internet addresses when the [redacted] for servicing. Private gateways [redacted] internet addresses, as desired.

Since the source or destination of a source-routed datagram may not have an internet address, it may be necessary to provide a return route for replies. This might be done by modifying the content of the original route to contain "back pointers" to intermediate destinations. Note that the local address of a private gateway in one network is usually different from its local address in the adjacent network.

Typically, a source would create a route which contains first the internet address of the host or gateway nearest to the desired destination. The next address in the route would be the local address of the destination. Figure 5 illustrates this notion. Host A.a wants to communicate with host Z. But Z is not attached to a formally recognized network.

To achieve its goal, host A.a can emit source-routed packets with the route: "B.y, Z." B.y identifies the host (private gateway) between net B and the new network as the first intermediate stop. The private gateway uses the "Z" information to deliver the datagram to the destination. When the datagram arrives, its route should contain "y, A.a" if the private gateway knows how to interpret A.a or "y, w, A.a" if the private gateway only knows about addresses local to network B.

Other Issues

The catenet model should provide for error messages originating within a network to be carried usefully back to the source. A global encoding of error messages or status messages is needed.

It is assumed that the gateway halves of a given network have a common status reporting, flow and congestion control mechanism. However, the halves on different nets may operate differently. There should be a defined interface between gateway halves which permits internet flow, congestion and error control to be exercised.

A gateway monitoring center (18) is postulated which can collect, correlate and display current gateway status. Such a center should not be required for the internet protocols to function, but could be used to manage an internet environment.

Accounting, accountability and access control procedures must also be defined for the global catenet.

Project Status

At the time of this writing (August 1978), the resources available to support the DARPA internet working research effort included:

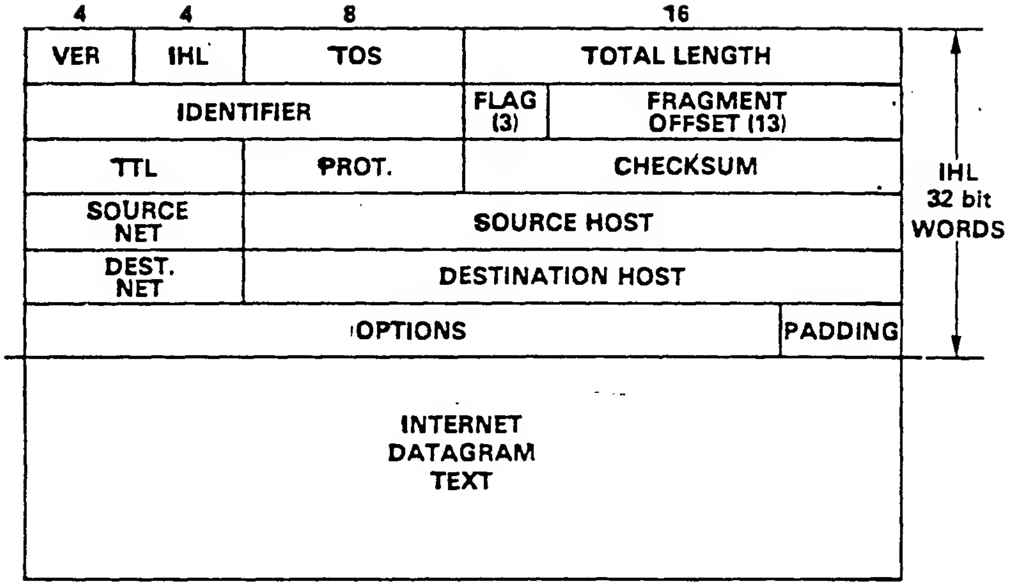
1. The ARPANET;
2. A major Packet Radio Network at SRI International in the San Francisco Bay area;
3. The Atlantic Packet Satellite Network;
4. The MIT LCS Network;
5. The Experimental Data Network at the U.S. Defense Communication Engineering Center; and
6. An interconnection to the UK Experimental Packet Switching System (EPSS).

A packet radio network is scheduled for installation in 1979 at Ft. Bragg, North Carolina for use in additional network experiments. This network will be connected to the ARPANET via gateway. All users of this network will depend on internet protocols to access resources in ARPANET.

The Catenet topology anticipated by June 1979 is illustrated in Figure 6. The internet protocols which will be used in the experimental research effort will include:

1. Internet datagrams (17);
2. Transmission control protocol (TCP) (19) - for reliable end/end transmission;
3. TELNET (20) - a virtual terminal protocol;
4. File Transfer Protocol (FTP) (20); and
5. Network Voice, Real-time and Internet Message System Protocols (under development).

The primary of the network-related research during 1979 will be to develop suitable flow control and alternate routing methods for Catenet gateways. Applications will concentrate on internet electronic message services including facsimile and command/control facilities such as mixed voice and graphics conferencing.



VER = VERSION TYPE

IHL = INTERNET HEADER LENGTH IN 32 bit WORDS

TOS = TYPE OF INTERNET SERVICE DESIRED
e.g. "LOW DELAY", "LOW COST", "HIGH BANDWIDTH",
"HIGH RELIABILITY", "DON'T DISCARD".

FLAG = CONTROL INDICATIONS SUCH AS
"OPTIONS PRESENT","MORE FRAGMENTS".

PROT = PROTOCOL IDENTIFIER (e.g. TCP, REAL-TIME)

INTERNET DATAGRAM FORMAT
FIGURE 3

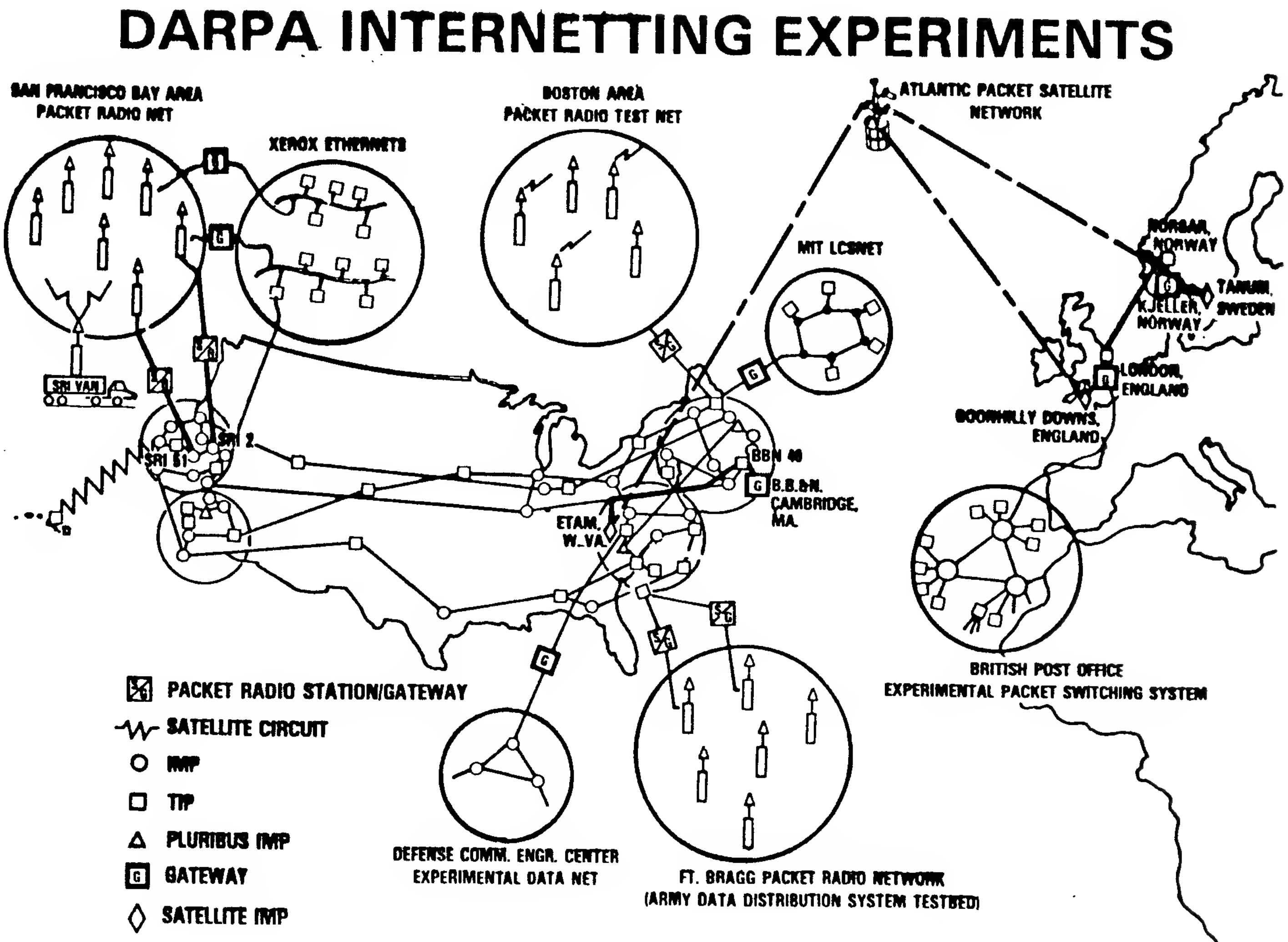

DARPA INTERNETTING EXPERIMENTS
SAN FRANCISCO BAY AREA PACKET RADIO NET
XEROX ETHERNETS
BOSTON AREA PACKET RADIO TEST NET
ATLANTIC PACKET SATELLITE NETWORK
MIT LCSNET
NORSAR, NORWAY
TANUM, SWEDEN
KJELLER, NORWAY
LONDON, ENGLAND
GOONHILLY DOWNS, ENGLAND
SRI VAN
SRI 2
SRI 51
BBN 40
B.B.&N. CAMBRIDGE, MA.
ETAM, W. VA.
BRITISH POST OFFICE EXPERIMENTAL PACKET SWITCHING SYSTEM
DEFENSE COMM. ENGR. CENTER EXPERIMENTAL DATA NET
FT. BRAGG PACKET RADIO NETWORK (ARMY DATA DISTRIBUTION SYSTEM TESTBED)
PACKET RADIO STATION/GATEWAY
SATELLITE CIRCUIT
IMP
TIP
PLURIBUS IMP
GATEWAY
SATELLITE IMP

FIGURE 6